



Attorney's Docket No.: 06998-022001

RECEIVED

DEC 20 1999

Group 2700

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Jonathan Hood Young et al. Art Unit : 2741
Serial No. : 08/825,534 Examiner : M. Lerner
Filed : March 28, 1997
Title : ERROR CORRECTION IN SPEECH RECOGNITION

Assistant Commissioner for Patents
Washington, D.C. 20231

APPEAL BRIEF

(1) Real Party in Interest

Dragon Systems, Inc., the assignee of this application, is the real party in interest.

(2) Related Appeals and Interferences

There are no related appeals or interferences.

(3) Status of Claims

Claims 1-6, 8-16, 24, 25, and 27-31 are pending in this application, with claims 1 and 25 being independent.

Claims 1-6 and 12 stand rejected as being anticipated by U.S. Patent No. 5,428,707 (Gould). Claims 8-11, 13-16, and 24 stand rejected as being obvious over Gould in view of U.S. Patent No. 5,027,406 (Roberts). Claims 25 and 27-31 stand rejected as being obvious over Roberts in view of U.S. Patent No. 5,677,990 (Junqua).

(4) Status of Amendments

Claims 17-23 have been cancelled subsequent to the final rejection. The amendment cancelling these claims accompanies the appeal brief.

(5) Summary of Invention

The invention relates to correcting errors in speech recognition.

A speech recognizer analyzes a user's speech to determine what the user said. The speech recognizer may recognize discrete words or phrases, which requires the user to pause between each discrete word or phrase, or it may recognize spoken words or phrases regardless of whether the user pauses between them. Page 1, lines 6-14.

The speech recognizer determines what the user said by finding acoustic models that best match the utterance, and identifying text that corresponds to those acoustic models. Acoustic models may represent a sound, a portion of a word (called a phoneme), silence, or various types of noise. The words or phrases corresponding to the best matching acoustic models are called recognition candidates. Page 1, line 24 to page 2, line2.

Speech recognizers have displayed a list of choices for each recognized word, and have permitted the user to correct a misrecognition by selecting a word from the list or by typing in the correct word. If the correct word was not on the list, the user could change the list either by typing the correct word, or by speaking words (for example, "alpha", "bravo") associated with the letters of the correct word. The user could also discard the entire list by saying "scratch that". Dictating a new word implied that the user accepted the previous recognition result. If, after dictating additional words, the user noticed an error, the user could say "oops" to bring up a numbered list of previously-recognized words from which the user could choose the number of the word in error. The user could then proceed as above and view a displayed list of choices for the word in error. Page 2, lines 7-19.

A speech recognizer may include speech recognition software running on a computer having input/output devices (such as a microphone, mouse, keyboard, and display), a processor, memory, and a sound card. The memory stores data and programs such as an operating system, an application program, and the speech recognition software. The microphone receives and conveys the user's speech to the sound card as an analog signal. The sound card uses an analog-to-digital converter and other components to transform the analog speech signal into a digital signal that can be analyzed by the processor. The processor, under control of the

operating system and the speech recognition software, processes the digital signal to identify utterances in the user's speech. Page 5, lines 16-26.

After processing the utterance, the speech recognizer provides a list of recognition candidates, where each recognition candidate has an associated score that indicates a probability that it corresponds to the user's speech. Recognition candidates may correspond to text or commands, and may include words, phrases, or sentences. Page 17, lines 4-16.

In generating a score for a recognition candidate, the speech recognizer uses acoustic scores for words of the candidate, a language model score that indicates a likelihood that words of the candidate are used together, and scores provided for each word of the candidate by a pre-filtering procedure. The scores provided by the pre-filtering procedure include a crude acoustic comparison and a language model score indicative of the likelihood that a word is used, independently of its context. Page 23, lines 16-22.

As mentioned above, recognition candidates may correspond to dictated text or commands. When the best-scoring recognition candidate corresponds to dictated text, the speech recognizer provides the text to an active application such as a word processor, and may also display the best-scoring candidate to the user through a graphical user interface. When the best-scoring recognition candidate is a command, the speech recognizer implements the command. Page 26, lines 5-13.

The speech recognizer has access to a complete dictation vocabulary, which includes the active vocabulary, and a backup dictionary. The backup dictionary may include user-specific backup vocabulary words, such as words created by the user while using the speech recognizer. The backup dictionary may further include system-wide backup vocabulary words. Page 27, lines 4-11.

When the speech recognizer makes a recognition error, the user may invoke an appropriate correction command to fix the error. The invention provides the correction commands "Spell That" and "Make That". When the speech recognizer determines that the best-

scoring recognition candidate includes a correction command, a correction dialog box is displayed and the speech recognizer enters a correction mode. Page 31, lines 14-27.

The user may choose either to speak a correct word or words using the "Make That" command, or to verbally spell a correct word or portion of a word using the "Spell That" command. When the user invokes the "Make That" command, the speech recognizer performs speech recognition on the utterance that includes the command and returns the results in the form of a list of ordered groups of recognition candidates. The speech recognizer may expand on the list of candidates by finding "confused pronunciation" matches for the phonemes following "Make That" in each of the recognition candidates. Confused pronunciation is based on the observation that phonemes having similar characteristics are commonly confused with one another. Page 33, lines 14-27.

When the user invokes the "Spell That" command, the speech recognizer recognizes the spelling of the word in the context of a constrained recognition that permits recognition of only letters. Recognition candidates are shown in the form of a list of ordered groups of letters, with each group being a probable recognition result for the word or portion of the word spelled by the user. The speech recognizer may find "confused spelling" matches for the groups of letters in the list of results. A confused spelling match is based on the observation that letters having similar pronunciations are often confused with one another. Page 34, lines 11-21.

In one general aspect, the invention features using a command, such as the "Make That" command, to correct incorrect text associated with recognition errors in computer-implemented speech recognition. Speech recognition is performed on an utterance to produce a recognition result for the utterance. A correction command is identified in one portion of the recognition result, and corrected text is identified from another portion of the recognition result. The correction command indicates that the portion of the recognition result includes a pronunciation of a word to be corrected. Page 3, lines 1-6.

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 5

Attorney's Docket No.: 06998-022001

In another general aspect, the invention features recognizing a spelling of a word using a command, such as the "Spell That" command, in computer-implemented speech recognition. Speech recognition is performed on an utterance to produce recognition results, and a spelling command is identified in the recognition results. The spelling command indicates that a portion of the utterance includes a spelling. The spelling is produced by using the recognition results and confused spelling matching to search a dictionary. In confused spelling matching, commonly-confused letters are treated as a single letter to identify the spelling corresponding to the portion of the utterance. Page 4, lines 13-17.

(6) Issues

Is the subject matter of claims 1-6 and 12 anticipated by Gould?

Would the subject matter of claims 8-11, 13-16 and 24 have been obvious over Gould in view of Roberts?

Would the subject matter of claims 25 and 27-31 have been obvious over Roberts in view of Junqua?

(7) Grouping of Claims

The claims do not stand or fall together. For the reasons presented in the Argument section below, the claims are believed to fall into five separately patentable groups, as follows: claims 1-3, 12, and 13; claims 8-11 and 14-16; claims 25-28; claims 29 and 30; and claim 31.

(8) Argument

**THE SUBJECT MATTER OF CLAIMS 1-6 AND 12 IS NOT
ANTICIPATED BY GOULD**

Independent claim 1 recites a method of correcting incorrect text associated with recognition errors in computer-implemented speech recognition. The method includes, among other steps, identifying a correction command in a recognition result for an utterance, and identifying corrected text from a portion of the recognition result for the utterance. Claim 1 further recites that the correction command indicates that the portion of the recognition result from which the corrected text is identified includes a pronunciation of a word to be corrected. Thus, the utterance on which speech recognition is performed includes a correction command and corrected text in the form of a pronunciation of a word to be corrected.

Appellant requests reversal of the rejection of claim 1 and the claims depending from claim 1 because Gould fails to describe or suggest processing an utterance that includes both a correction command and corrected text in the form of a pronunciation of a word to be corrected. The Examiner argues that Gould describes such an utterance with respect to the correction commands “Choose-N” and “Scratch-That”. As discussed below, the Examiner is incorrect.

Neither of the “Choose-N” and “Scratch That” commands includes corrected text in the form of a pronunciation of a word to be corrected. The “Choose-N” command indicates that the user wishes to select a “specifically numbered word shown on a currently displayed choice menu as the intended word for the utterance represented by that choice window....” See Gould at col. 9, lines 1-4. Thus, to select the third word on a displayed list, a user would say “Choose-3”. Therefore, the “Choose-N” command uses a number to select corrected text from a displayed list, and does not use corrected text in the form of a pronunciation of a word to be corrected.

Similarly, the "Scratch That" command indicates that the user wishes to select "none of the words displayed in the current choice window, including the first choice word...." See Gould at col. 11, lines 55-60. Therefore, the "Scratch That" command includes only the words "Scratch" and "that" and does not include corrected text in the form of a pronunciation of a word to be corrected.

Apparently, in view of these features of Gould, the Examiner uses an overly broad construction of the term "utterance", so as to include, for example, utterance of the word "very" (that is, a first utterance), followed by the later utterance of the command "Choose-3" (that is, a second utterance), within the confines of a single utterance. At page 10 of the final action, the Examiner explains that:

Broadly speaking, "an utterance" may be construed to be continuous dictation of original text as well as correction command with associated corrected text.

and further writes:

In the example of Figure 49, a user first says "very," and then sometime later says "choose 3." The utterance therefore comprises "very...choose 3."

Initially, Appellant notes that the Examiner's treatment of two utterances as a single utterance contradicts both the definition of an utterance set forth in the application, and the well-understood definition of an utterance as a term in the art, as evidenced by Gould. The application at, for example, pages 5-6, notes that utterances are "separated from one another by a pause having a sufficiently large predetermined duration (e.g., 16-250 milliseconds)" and that "[e]ach utterance may include one or more words of the user's speech." Gould defines an utterance similarly: "the utterance to be recognized will normally be proceeded [sic, preceded]

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 8

Attorney's Docket No.: 06998-022001

and followed by silence in a discreet [sic, discrete] utterance recognizer in which words to be recognized are to be spoken separately" (col. 2, lines 3-6).

The Examiner, in the Advisory Action, indicates that Gould's definition of an utterance, as quoted above at col. 2, lines 3-6 of Gould, refers only to discrete speech recognition systems. Even assuming for the sake of argument that the Examiner is correct, the fact remains that this is the only definition of an utterance provided by Gould. Accordingly, nothing in Gould would have led one of ordinary skill in the art to adopt the overly broad definition asserted by the Examiner.

Moreover, in the example noted by the Examiner at page 9 of the final action, and relied on by the Examiner as support for the broad construction of "utterance", Gould treats "very" as an utterance separate from the utterance of "Choose-3". As discussed by Gould at col. 7, line 64 to col. 9, line 25, and illustrated in Fig. 5, an utterance may be a word, which results in removal of the previous choice window, simulated typing of the best scoring word, and creating of a new choice window (see steps 218-224 of Fig. 5). An utterance also may be a choice command, which results in replacement of the best scoring word with the chosen word and removal of the choice window (see steps 216 and 226-234 of Fig. 5). As noted by Gould at col. 27, lines 11-15 and illustrated in Fig. 46, saying the word "very" causes the incorrect word "vary" to be displayed along with a choice window, thus establishing "very" as a first utterance (that is, an utterance of a word). When the user later says "Choose-3", this is treated as a second, separate utterance (that is, an utterance of a choice command). See Gould at col. 27, lines 41-46.

Indeed, the Examiner, apparently contradicting his own overly broad definition of an utterance given above, supports the Appellant's assertion that "very" and "Choose-3" are separate utterances (page 10 of final action):

For the utterance "very," DragonDictate produces the (mis)recognition result "vary." For the utterance "choose 3," a Choose-N correction command is identified, and the program automatically corrects the last recognition result ("vary")...

Finally, even if "very" and "Choose-3" could somehow be said to constitute parts of a single utterance, "very" could not be said to constitute "corrected text," since, at the time "very" is spoken, there is no error to be corrected. Rather, the misrecognition of "very" as "vary" is the error that requires correction through use of the "Choose-3" command. For each of these reasons, claim 1 is not anticipated by Gould's "Choose-N" command.

The Examiner also refers to Gould's "Scratch-That" command as somehow identifying corrected text by stating that "identification and removal of incorrect text corrects text in some cases." This is incorrect for at least the following reasons. The process of correcting text, as embodied in the "Scratch-That" command, and pointed to by the Examiner, is not equivalent to identifying corrected text, as recited in claim 1. Identifying corrected text is just one way of correcting text. Moreover, the process of correcting text does not require the identifying of corrected text. Rather, as evidenced by the "Scratch-That" command, the process of correcting text may simply include the removal of incorrectly recognized text. Therefore, the "Scratch-That" command does not identify corrected text but simply removes incorrectly recognized text. For these reasons, claim 1 is not anticipated by Gould's "Scratch-That" command.

The Examiner also indicates that Gould's adaptive training subroutine somehow indicates that the portion of the recognition result for an utterance of a correction command includes a pronunciation of a word to be corrected, as recited in claim 1. However, the adaptive training subroutine (as described in Fig. 12 and called in the routine of Fig. 5 of Gould) is only used to improve word models for a vocabulary. See Gould at col. 10, lines 3-16. The Examiner states at page 10 of the final office action that the last token recognized is automatically stored for the entry in the OOPS buffer to correct the word if the word turns out to have been misrecognized. However, the stored token is never used by the commands in Gould to correct the word. The stored token is used initially by the system to determine the words in the choice list. Then the stored token is used after a word has been correctly labeled (using, for example, the Choose-N command) to perform adaptive training on word models. Because the token in

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 10

Attorney's Docket No.: 06998-022001

Gould's system is used in the adaptive training subroutine only after the user has selected the correct word using a command such as Choose-N, the token cannot be considered to be corrected text including a "pronunciation of a word to be corrected," as recited in claim 1.

Accordingly, the rejection of claim 1 should be reversed, as should the rejection of the claims depending from claim 1.

**THE SUBJECT MATTER OF CLAIMS 8-11, 13-16 AND 24
WOULD NOT HAVE BEEN OBVIOUS OVER GOULD IN
VIEW OF ROBERTS**

Claims 8-11, 13-16, and 24 all depend from claim 1. Appellant requests reversal of the rejection of these claims because Roberts does not remedy the failure of Gould to describe or suggest the subject matter of claim 1. In particular, neither Gould, Roberts, nor the combination of the two describes or suggests processing an utterance that includes both a correction command and corrected text in the form of a pronunciation of a word to be corrected. Roberts' correction commands are "start_comletter", which is used to edit mistakenly recognized words, and "backspace", which deletes the last letter added to a string of letters that spell a word. As has been conceded by the Examiner, these correction commands do not include a pronunciation of a word to be corrected, as recited in claim 1. Therefore, since both Gould and Roberts lack this feature of claim 1, any possible combination of Gould and Roberts would fail to describe or suggest this feature, the rejection of claims 8-11, 13-16, and 24 should be reversed.

The rejection of claim 8 and the claims depending from claim 8 should be reversed for the additional reason that neither Gould, Roberts, nor any combination of the two describes or suggests identifying corrected text using confused pronunciation matching to identify text corresponding to the pronunciation as recited in claim 8. In confused pronunciation matching, commonly-confused sounds are treated as a single sound. The Examiner concedes that Gould does not disclose or suggest identifying corrected text using confused pronunciation matching. The Examiner then attempts to use Roberts to make up for this shortcoming of Gould

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 11

Attorney's Docket No.: 06998-022001

by stating that the “phonetic dictionary 500a of *Roberts et al.* is used with correction commands (column 21, lines 41 to 56) to perform ‘confused pronunciation matching’.” This is simply incorrect because Roberts never describes confused pronunciation matching. Rather, Roberts simply states that the backup vocabulary need not be restricted when the backup dictionary includes phonemic pronunciations of possible words (referred to as a phonemic dictionary). Roberts defines such a phonemic dictionary as including word entries that have associated spellings in phonemic symbols. Roberts further explains that, if there are too many choices for an utterance when using the phonemic dictionary, then “the language model can be used to further limit this vocabulary to a number of words which can be recognized sufficiently promptly.” Using the example given by Roberts, the “schwa” sound occurs in both the words “about” and “above” before the letter “b”. Roberts then builds up a phonemic model of the “schwa” sound that takes into account the “b” sound that often follows it. Roberts never suggests that another phoneme is ever confused with either the “schwa” sound or the “b” sound. This is because Roberts is not describing confused pronunciation matching, in which commonly-confused sounds are treated as a single sound. Accordingly, the rejection of claim 8 and all claims depending from claim 8 should be reversed for this additional reason.

**THE SUBJECT MATTER OF CLAIMS 25 AND 27-31
WOULD NOT HAVE BEEN OBVIOUS OVER ROBERTS IN
VIEW OF JUNQUA**

Independent claim 25 recites a method for recognizing a spelling of a word in computer-implemented speech recognition. The method includes performing speech recognition on an utterance to produce recognition results for the utterance and identifying a spelling command in the recognition results. The spelling command indicates that a portion of the utterance includes a spelling. The method further includes producing the spelling by searching a dictionary using the recognition results. Producing the spelling includes using confused spelling

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 12

Attorney's Docket No.: 06998-022001

matching. In confused spelling matching, commonly-confused letters are treated as a single letter to identify the spelling corresponding to the portion of the utterance.

Appellant requests reversal of the rejection of claim 25 and the claims depending from claim 25 because there would have been no motivation to combine Roberts with Junqua in the manner suggested by the Examiner. Initially, Appellant notes that, as apparently conceded by the Examiner, Roberts fails to describe or suggest using confused spelling matching and instead uses a phonetic alphabet. The Examiner argues that Roberts somehow suggests that spelling may be confusingly similar and that this supposed suggestion by Roberts is equivalent to confused spelling matching. However, this statement is wrong for at least the following reasons. First, Roberts does not suggest that letters may be confusingly similar. For example, at col. 20, lines 12-19, Roberts states that:

commands for adding individual letters to the STARTSTRING [may] be changed from "starts_alpha" through "starts_zulu" to the more quickly said "alpha" through "zulu". This change is made possible by the fact that the EDITMODE's restricted vocabulary does not contain text words which are the same as, or confusingly similar to, the words of the communications alphabet.

Therefore, Roberts removes words from the restricted vocabulary that sound confusingly similar to words from the communications alphabet, and has nothing to do with processing of similarly-sounding letters. For example, Roberts might remove the word "voodoo" from the restricted vocabulary because it sounds similar to the word "zulu" in the communications alphabet. Moreover, even assuming for the sake of argument that Roberts does suggest that letters may be confused with other letters, this still does not describe or suggest confused spelling matching, in which commonly-confused letters are treated as a single letter.

There would have been no motivation to employ Junqua's confused spelling matching in Roberts' system because neither Roberts nor Junqua describes or suggests that confused spelling matching improves recognition accuracy in a speech recognizer that otherwise

uses a phonetic alphabet. The Examiner states that Junqua's confused spelling matching makes letter commands or a phonetic alphabet unnecessary: "The point of the teaching of *Junqua*, however, is that confused spelling matching makes letter commands or a phonetic alphabet unnecessary." In particular, the Examiner uses an example of a misrecognition of the word "invention" as "inversion" to give motivation for using Junqua in the system of Roberts. According to the Examiner, Roberts' "starts_eye..." command that corrects this misrecognition produces a much longer list of recognition candidates than the Examiner's suggested command "starts_i...n...v...e...n...t...."

The Examiner then states that Junqua suggests at col. 1, lines 50-67 that confused spelling matching would further decrease the response time by producing an even shorter list of candidates. This is incorrect. Since confused spelling matching permits a single letter to represent multiple letters, confused spelling matching will actually produce a longer list of candidates.

In addition to this fundamental flaw, the Examiner's argument includes a number of other problems. First, contrary to the Examiner's assertion, Junqua, at col. 1, lines 50-67, in no way suggests that confused spelling matching decreases the response time by producing a shorter list of candidates. Rather, that passage states that, although reasonable accuracy may be obtained using a fixed list such as a telephone directory to constrain spelling recognition, response time will increase if the size of the list increases: "while reasonable recognition accuracy can be obtained using a conventional speech recognizer constrained on the sequence of possible letters by the known list, response time increases quite dramatically as the size of the list or dictionary increases." Therefore, according to Junqua, a system that does not use a large list (that is, a large fixed list or dictionary) decreases response time.

This is why Junqua breaks up the speech recognition process into steps. As Junqua states at col. 1, lines 62-65 -- "To attain optimally short response time, the processes are performed first without costly constraints and thereafter with costly constraints, if needed, after the number of word candidates is low." Thus, confused spelling is used only after the first pass

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 14

Attorney's Docket No.: 06998-022001

produces an N-best candidate list. Junqua explains that a neural network discriminator distinguishes

between confusable letters (such as the letters J and K). The neural network is applied to confusable subsets. The first pass of the HMM recognizer produces a sequence of letters (one sequence for each of the N-best). If one of these letters belongs to a confusable subset, the neural network discriminator is launched

Junqua does this because confused spelling matching, when used on an entire name dictionary, tends to increase response time. Therefore, Junqua explains that confused spelling matching is performed on a "dynamic grammar" as opposed to the entire name dictionary: "Having selected the N-best candidates, these candidates are used to build a dynamic grammar. With this dynamic grammar in place, the sequence of letters uttered into the microphonic transducer is then processed a second time through a speech recognizer."

Furthermore, Junqua's speech recognition system is said to gain the highest recognition accuracy not from confused spelling matching (as the Examiner suggests), but from the highly constrained recognition at the fourth pass. Junqua states at col. 8, lines 56-63 that "by reserving the highly constrained (high detail) recognizer for the 4th pass, the system can recognize continuously spelled names without a great deal of computational overhead." Thus, Junqua in no way suggests that confused spelling matching would be useful in a system such as Roberts'.

In addition, contrary to the Examiner's assertion at page 12 of the final action, Fig. 15 of Roberts has nothing to do with speaking a "starts_invent" command. Fig. 15 of Roberts does not relate to speech recognition of letters, but instead relates to processing of typed letters. Specifically, Fig. 15 relates to a condition in which a user first types the letter "i" into the system, thus restricting the candidate list to words that begin with the letter "i". The user then types "n", "v", and "e". Roberts explains at col. 24, lines 32-63, that in Fig. 15, "the user types the letter

command 'i' to indicate that the utterance to be recognized begins with that letter" and "the user next types the letters 'n', 'v', 'e' in successive passes"

Furthermore, although the Examiner claims there is motivation for using confused spelling matching in the system of Roberts because confused spelling matching somehow improves recognition accuracy, neither Junqua nor Roberts states or implies that using confused spelling matching would improve recognition accuracy in a speech recognition system that would otherwise use a phonetic alphabet (as used in Roberts). To the contrary, Junqua implies at col. 1, lines 42-49 that using a phonetic alphabet actually eliminates the need for confused spelling matching because using a phonetic alphabet improves recognition accuracy: "Recognition of spoken letters is even difficult for humans . . . This is why radio telephone operators are trained to use a phonetic alphabet, A-Alpha, B-Baker, C-Charlie, etc., when communicating over a noisy channel." Similarly, Roberts states at col. 19, line 57 to col. 20, line 19 that the phonetic alphabet commands are used with a restricted vocabulary (or list) in the EDITMODE. This is done in Roberts because a large vocabulary or list is not needed when using the phonetic alphabet because the phonetic alphabet produces increased recognition accuracy.

Roberts fails to describe or suggest using confused spelling matching to identify the spelling corresponding to the portion of the utterance. Moreover, for the reasons noted above, there would have been no reason to employ the confused spelling matching system of Junqua in Roberts' system. As noted above, if a special phonetic alphabet is used (such as the communications or phonetic alphabet of Roberts), spelling recognition accuracy in a speech recognizer is greatly improved. With this improved recognition accuracy, there is no need to employ confused spelling matching in Roberts' speech recognition system. Accordingly, one of ordinary skill in the art would have had no motivation to combine Roberts and Junqua in the manner suggested by the Examiner.

Accordingly, the rejection of claim 25 should be reversed, as should the rejection of the claims depending from claim 25.

Claim 29 recites that the method of claim 25 further includes generating a list of confused spelling matches and identifying the text corresponding to the spelling as a selection from the list of confused spelling matches. As discussed above, and as conceded by the Examiner at pages 7-8, Roberts does not "disclose 'commonly-confused letters are treated as a single letter to identify the spelling corresponding to the portion of the utterance.'" Therefore, Roberts cannot generate a list of confused spelling matches, and neither can Roberts identify text from the list of confused spelling matches, as recited in claim 29. The Examiner suggests that the list of words displayed in Figs. 10-24 of Roberts and labeled as 701, 702, corresponds to a list of confused spelling matches. However, as explained in Roberts at col. 21, lines 17-37, the displayed list in Roberts is simply a list of the best recognition choices:

Step 274 displays the best choices from that recognition in the active window 701 a second window, called the dictionary window 702, is displayed directly below the window 701. The dictionary window contains enough of the backup words selected from the backup dictionary by step 268, to bring the total number of choices presented in both the active window 701 and the dictionary window 702 up to a total of nine words.

Therefore, Roberts does not provide a list of confused spelling matches. Accordingly, the rejection of claim 29, and claim 30, which depends from claim 29, should be reversed for this additional reason.

Claim 31 recites that confused spelling matching includes using characteristics of a speaker's pronunciation to search the dictionary for confused spelling matches. As detailed above, Roberts does not describe or suggest using confused spelling matching. Moreover, Junqua fails to describe or suggest using characteristics of the speaker's pronunciation during confused spelling matching, as recited in claim 31. The Examiner states that "*Junqua* discloses that confused spelling matching uses vowel sounds, fricatives, affricatives, plosives and nasals ("characteristics of a speaker's pronunciation") to provide distinguishing features between confusable letters." *Junqua* does use the above-mentioned portions of an utterance to provide

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 17

Attorney's Docket No.: 06998-022001

distinguishing features between confusable letters. However, Junqua never suggests using characteristics of a speaker's pronunciation to distinguish features between confusable letters. This is so because the characteristics mentioned above are not characteristics of a speaker's pronunciation. Rather, these characteristics are characteristics of the spoken language. For example the "p" in the word "top" is pronounced as a plosive regardless of who is speaking the word. However, if a speaker pronounces the word "top" as -täp-, and not as the correct pronunciation of "top", which is -töp-, then a characteristic of the speaker's pronunciation would include pronouncing -ö- as -ä-.

Because the characteristics used in Junqua are characteristics of speech and not characteristics of the speaker's pronunciation, neither Roberts nor Junqua describes or suggests using characteristics of the speaker's pronunciation, as recited in claim 31. Accordingly, the rejection of claim 31 should be reversed.

Conclusion

For the foregoing reasons, the rejections should be reversed.

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 18

Attorney's Docket No.: 06998-022001

This appeal brief is being filed within three months from September 22, 1999, the filing date of the Notice of Appeal. Attached hereto is a Petition for Extension of Time for one month extending to and including December 22, 1999. Please apply the fee of \$300.00 for this brief under 37 C.F.R. §1.17(c) and the fee of \$110.00 for the extension of time under 37 C.F.R. §1.136 to Deposit Account No. 06-1050, Reference No. 06998-022001. If there are any charges not covered, or any credits, please apply them to Deposit Account No. 06-1050, Reference No. 06998-022001.

Respectfully submitted,

Date: 12/16/99


John F. Hayden
Reg. No. 37,640

JFH/DZD/adt

Fish & Richardson P.C.
601 Thirteenth Street, NW
Washington, DC 20005
Telephone: (202) 783-5070
Facsimile: (202) 783-2331

40017536.wpd

(9) Appendix - All claims as currently pending.

1. A method of correcting incorrect text associated with recognition errors in computer-implemented speech recognition, comprising:

 performing speech recognition on an utterance to produce a recognition result for the utterance;

 identifying a correction command in the recognition result for the utterance; and
 identifying corrected text from a portion of the recognition result for the utterance,
 wherein the correction command indicates that the portion of the recognition
 result comprises a pronunciation of a word to be corrected.

2. The method of claim 1, further comprising replacing previously-generated incorrect text with the corrected text.

3. The method of claim 1, wherein the step of identifying corrected text includes searching a dictionary using the portion of the recognition result.

4. The method of claim 1, wherein the step of identifying corrected text comprises identifying corrected text from a portion of the recognition result for the utterance and from a recognition result for a second utterance.

5. The method of claim 4, wherein the second utterance precedes the utterance.

6. The method of claim 4, wherein the second utterance follows the utterance.

8. The method of claim 1, wherein the step of identifying corrected text comprises using confused pronunciation matching, in which commonly-confused sounds are treated as a single sound, to identify text corresponding to the pronunciation.

9. The method of claim 8, wherein the confused pronunciation matching comprises using the pronunciation to search a confused pronunciation dictionary.

10. The method of claim 8, wherein the confused pronunciation matching comprises using the pronunciation to search a pronunciation dictionary for confused pronunciation matches.

11. The method of claim 8, wherein the confused pronunciation matching comprises using a phonetic tree to search a pronunciation dictionary.

12. The method of claim 2, further comprising automatically selecting the previously-generated incorrect text to be replaced.

13. The method of claim 12, wherein the step of automatically selecting comprising re-recognizing previously-recognized speech corresponding to the previously-generated incorrect text using the corrected text.

14. The method of claim 8, further comprising generating a list of confused pronunciation matches and identifying the corrected text as a selection from the list of confused pronunciation matches.

15. The method of claim 14, further comprising using the list of confused pronunciation matches to re-recognize previously-recognized speech so as to determine the corrected text.

16. The method of claim 14, further comprising displaying text corresponding to the list of confused pronunciation matches to a user and obtaining the selection from the user.

24. The method of claim 1, the method further comprises:
using an active vocabulary when performing speech recognition,
using a backup dictionary when identifying the corrected text, and
if the active vocabulary does not contain the corrected text, adding the corrected text to the active vocabulary.

25. A method for recognizing a spelling of a word in computer-implemented speech recognition, comprising:
performing speech recognition on an utterance to produce recognition results;
identifying a spelling command in the recognition results, wherein the spelling command indicates that a portion of the utterance comprises a spelling;
producing the spelling by searching a dictionary using the recognition results,
wherein producing the spelling comprises using confused spelling matching, in which commonly-confused letters are treated as a single letter, to identify the spelling corresponding to the portion of the utterance.

27. The method of claim 25, wherein the dictionary is a confused spelling dictionary and the confused spelling matching comprises using the spelling to search the confused spelling dictionary.

Applicant : Jonathan Hood Young et al.
Serial No. : 08/825,534
Filed : March 28, 1997
Page : 22

Attorney's Docket No.: 06998-022001

28. The method of claim 25, wherein the dictionary is a spelling dictionary and the confused spelling matching comprises using the spelling to search the spelling dictionary for confused spelling matches.

29. The method of claim 25, further comprising generating a list of confused spelling matches and identifying the text corresponding to the spelling as a selection from the list of confused spelling matches.

30. The method of claim 29, further comprising displaying the list of confused spelling matches to a user and obtaining the selection from the user.

31. The method of claim 25, wherein confused spelling matching comprises using characteristics of a speaker's pronunciation to search the dictionary for confused spelling matches.